MASTER'S THESIS

Lecturers' Beliefs and Needs concerning Embedded Assessment in Online Higher Education:an Exploratory Study.

Snoeys, Liesbet

Award date:
2020

Lecturers’ Beliefs and Needs concerning Embedded Assessment in Online Higher Education: An Exploratory Study

De Opvattingen en de Behoeften van Docenten aangaande Embedded Assessment in het Online Hoger Onderwijs: Een Verkennende Studie

L. Snoeys

Master Onderwijswetenschappen
Open Universiteit

Datum: 25 januari 2020
Begeleiding: Dr. J. Janssen

Table of content

Lecturers' Beliefs and Needs concerning Embedded Assessment in Online Higher Education: an Exploratory Study

L. Snoeys

Summary

Background. Recent development towards data driven education have led the Open University (OUNL) to explore and elaborate technical possibilities of its digital learning environment and wider data infrastructure. Complementary to these technical explorations, this study investigates educational affordances of data driven education, more particularly, embedded assessment: continuous, nearly unnoticed data collection while learners execute learning activities. This kind of assessment is considered to raise assessment validity, especially with regard to highly complex skills, and has a combined formative-summative purpose.

Aim. This thesis research is a first exploration of OUNL's lecturers' beliefs and needs considering embedded assessment in their own educational practice. Based on this, tentative design principles for embedded assessment in the particular educational context are formulated.

Participants, procedure, design. Nine lecturers of the OUNL master of Educational Sciences, varying in age (M = 44.67, SD = 12.18) and professional experience (less than five to more than 20 years), consented to participate in individual open-ended interviews. This qualitative exploratory study is carried out as a first step (systematic analysis) within a design-based research design.

Measures. The materials used during the interviews include an interview protocol and an introductory video explaining the concept of embedded assessment to participants. Before use, these materials were pilot tested and refined.

Results. Participants' general reflections on embedded assessment were positive but with some reservations regarding the use of embedded assessment for complex skills as well as for summative purposes. Embedded assessment was expected to increase the availability and effectiveness of (automated) feedback, to facilitate agile and differentiated instructional adaptations, and to foster more reliable summative decision making. However, lecturers expressed multiple concerns regarding assessment validity. Anticipated effects of embedded assessment implementation on student experience and self-regulation and on lecturer work load were not without ambiguity. All lecturers perceived concrete opportunities for future use of embedded assessment in their own courses, mainly in support of formative assessment purposes. Perceived opportunities focus on specific learning objectives reflected in products the student creates as well as on student self-regulation and collaborative learning.

Conclusion. Based on lecturers' views and suggestions, tentative design principles for embedded assessment are formulated. Considering possible future implementation of embedded assessment in this particular educational program, further requirement elicitation and thorough research regarding underlying learning progressions are necessary. With regard to the general exploration of lecturers' beliefs and needs regarding embedded assessment, future research could focus on different educational programs and universities. The developed embedded assessment framework is found to be useful to introduce the concept to participants, although explicating the possibility to provide embedded assessment information directly to students seems necessary. The frameworks clarity to other potential user groups should be tested before use.

Keywords: embedded assessment – lecturers' beliefs - needs assessment – qualitative research - design-based research

De Opvattingen en de Behoeften van Docenten aangaande Embedded Assessment in het Online Hoger Onderwijs: een Verkennende Studie

L. Snoeys

Samenvatting

Achtergrond. Recente ontwikkelingen richting *data driven education* leidden bij Open Universiteit (OUNL) tot het verkennen en het uitbreiden van de technische mogelijkheden van de digitale leeromgeving en de bredere datainfrastructuur. In aanvulling hierop richt dit onderzoek zich op hoe data driven education de onderwijspraktijk zou kunnen versterken in de vorm van *embedded assessment*: het continu, bijna onopgemerkt verzamelen van data tijdens de uitvoering van leeractiviteiten. Embedded assessment wordt gezien als beloftevol voor het verhogen van de assessmentvaliditeit, in het bijzonder voor complexe vaardigheden, en heeft een gecombineerd formatief-summatief doel.

Doel van dit thesisonderzoek is een eerste verkenning van de opvattingen en de behoeften van OUNL-docenten aangaande embedded assessment in hun eigen onderwijspraktijk, als basis voor het formuleren van (voorlopige) ontwerpprincipes voor embedded assessment in deze specifieke onderwijscontext.

Deelnemers, procedure, onderzoeksontwerp. Negen OUNL-docenten van de master Onderwijswetenschappen, met uiteenlopende leeftijden (M = 44.67, SD = 12.18) en professionele ervaring (minder dan vijf tot meer dan twintig jaar), namen deel aan individuele *open-ended* interviews. Dit kwalitatieve verkennende onderzoek vormt een eerste stap (systematische analyse) in een ontwerpgericht onderzoeksproces.

Meetinstrumenten. Het gebruikte materiaal bestaat uit een interviewprotocol en een inleidende video om het concept embedded assessment aan de participanten toe te lichten. Een *pilot test* voor gebruik maakte deel uit van de ontwikkeling.

Resultaten. De participanten drukten zich in het algemeen positief uit over embedded assessment, maar uitten terughoudendheid wat betreft het gebruik in functie van complexe vaardigheden en het nemen van summatieve beslissingen. Ze verwachtten door embedded assessment meer beschikbaarheid en effectiviteit van (automatische) feedback, facilitatie van snelle en gedifferentieerde aanpassingen van de instructie en meer betrouwbare summatieve beslissingen. Wat betreft de validiteit van embedded assessment formuleerden de participanten verschillende bezorgdheden. De veronderstelde effecten van de implementatie op de ervaringen en de zelfregulatie van de student en de werkbelasting voor de docent zijn ambigu. Elke deelnemer beschreef concrete mogelijkheden voor het toekomstige gebruik van embedded assessment in de eigen opleidingsonderdelen, voornamelijk in

functie van formatief assessment. De genoemde mogelijkheden omvatten zowel specifieke leerdoelen die tot uiting komen in producten die de student maakt, als zelfregulatie en samenwerkend leren.

Conclusie. De antwoorden van de docenten leiden tot het formuleren van een aantal voorlopige ontwerpprincipes voor embedded assessment. In het kader van het mogelijk toekomstig implementeren van embedded assessmentontwerp binnen deze opleiding is verdere analyse van vereisten noodzakelijk, net als het grondig onderzoeken van onderliggende *learning progressions*. Wat de algemene verkenning van de opvattingen en de behoeften van lectoren aangaande embedded assessment betreft, kan toekomstig onderzoek zich richten op docenten aan andere opleidingen en universiteiten. Het ontwikkelde raamwerk voor embedded assessment blijkt nuttig om het concept aan de participanten toe te lichten, hoewel het nodig lijkt om te expliciteren dat embedded assessment informatie ook rechtstreeks aan de student kan worden bezorgen. De duidelijkheid van het raamwerk voor andere potentiële gebruikersgroepen moet voor gebruik worden nagegaan.

Keywords: embedded assessment – opvattingen van docenten – behoeftenanalyse – kwalitatief onderzoek – ontwerpgericht onderzoek

Introduction

Traditionally, teachers base summative judgement of students on explicit assessments (Redecker & Johannessen, 2013). Formative evaluation, aimed at providing feedback or adapting instruction, is based on additional intermediate testing and instantaneous observations (Black & Wiliam, 1998; Haug & Ødegaard, 2015; Heritage, 2007). Recently, authors put forward to systematically integrate formative and summative assessment functions (Black & Wiliam, 2018). Concurrently, various authors propose a shift in the data collection process towards largely unnoticed collection of data during learning: *embedded assessment* (Farrell & Rushby, 2016; Redecker & Johannessen, 2013; Shute, Leighton, Jang, & Chu, 2016).

Embedded assessment provides a more continuous mapping of the student's learning process and performance in relation to the learning objectives, thus facilitating targeted feedback and feedforward, as well as agile adjustment of instruction when required (Redecker & Johannessen, 2013; Shute & Kim, 2013; Shute, Ventura, Bauer, & Zapata-Rivera, 2009). Redecker and Johannessen (2013) consider this evolution towards embedded assessment a necessary one in order to asses and facilitate the development of complex skills: embedded assessment allows mapping these skills in multiple different, complex, authentic situations. Compared to assessing a limited representation of competencies in defined tasks during one single moment of evaluation (Black & Wiliam, 2018), embedded assessment raises assessment validity.

Publications on embedded assessment – still an emerging field – focus mainly on global conceptual frameworks (e.g. Shute, et al., 2016; Farrell & Rushby, 2016). Subsequent design and development of effective embedded assessment practices requires a thorough analysis of the design context (Edelson, 2002). For embedded assessment this design context is determined by:

- characteristics of learning objectives and learning activities;
- the digital learning environment in which learning activities are integrated, and the data collected in connection with these activities; and
- user (in the context of this study: teacher) characteristics.

At the Open University of the Netherlands (OUNL), students acquire complex skills through online active learning (Open Universiteit, 2019a) in a digital learning environment largely developed in-house. Various ongoing projects investigate the possibilities of collecting and combining data generated by the learning environment and student information system. Not yet investigated is how lecturers think of using digital learning environment data for embedded assessment. Determining the conceptions and needs of this user group however is a crucial starting point (Coburn & Turner, 2011; Flower, 1985; Zapata-Rivera & Katz, 2014; Zenisky & Hambleton, 2012) for formulating design principles for embedded assessment in this specific context.

Embedded Assessment

In publications embedded assessment refers to different conceptualizations of assessment integrated in courses or curricula (e.g. Cummings, Maddux, & Richmond, 2008; Furtak, et al., 2008; Gerretson & Golson, 2004; Kerby & Romine, 2009; Park, Seo, You, & Song, 2016; Pike, 2014; Shavelson, et al., 2008). In light of this research we define embedded assessment as continuous, nearly unnoticed data collection while learners execute learning activities in the context of a course, a series of courses or an entire educational program, aimed at evaluating and stimulating the development of students' competences (Farrell & Rushby, 2016; Johnson-Glenberg, 2010; Shute, et al., 2016; Wilson & Sloane, 2000).

Embedded assessment has a combined formative-summative purpose (Shute, et al., 2016). It provides insight in the student's performance (result) and learning process (process) relative to the learning objectives, which enables judgement (*assessment of learning*), targeted feedback and -forward and/or agile adjustment of instruction if needed (*assessment for learning*) (Redecker & Johannessen, 2013; Shute & Kim, 2013; Shute, et al., 2009; Webb, Gibson, Forkosh-Baruch, 2013).

Embedded Assessment Data

Embedded assessment data are continuously collected data related to the student's performance and learning process. Collection spans a variety of learning activities, integrated in educational technology, over a longer period of time. Data are specific observable aspects of the knowledge, skills and attitudes that are to be acquired (Shute, et al., 2016). Consequently, the design of embedded assessment is determined by the learning activities a student carries out and the digital learning environment in which these activities are integrated, including its data collection affordances (Figure 1).

Various authors describe examples of embedded assessment data. Libin, et al. (2010), for instance, describe a digital learning environment in which short video cases are followed by multiple-choice questions, offering different possible reactions to the displayed professional situation. Learners responses throughout subsequent situations are registered. Ridgway & McCusker (2008) mention key strokes registration during ICT-focused learning activities. Also, they suggest that – following informed consent – spyware can be installed on student devices in order to register the student's online actions during literature search. Visited URLs and/or typed search terms could be among the data collected. As a final example, discussing artificial intelligence in education, Luckin (2018) describes automated analysis of video recordings of students jointly approaching a task, in order to determine the quality of student cooperation.

Data about the student's subsequent actions during learning activities and about the results related to these actions inform about the student's learning process (Thille & Zimmaro, 2017). Based on data analysis per constituent learning goal and its related evaluation criteria, progress and recurring strengths and errors can be determined (Ellis, 2013). By combining data from multiple learning activities and different subsequent moments, individual variations in learning processes can be registered (Rose & Fisher, 2011).

Redecker and Johannessen (2013) frame embedded assessment as a specific application of learning analytics, “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Siemens & Gasevic, 2012, p. 1). Embedded assessment data about observable aspects of knowledge, skills and attitudes during concrete learning activities aimed at formative and summative decisions and related actions (Shute, et al., 2016) can be considered a subset of learning analytics data. Other types of – frequently used - learning analytics data are registered events aimed at more broadly mapping the student’s learning behaviour and results. Examples are time spent on page or task (Tempelaar, Rienties, & Giesbers, 2015; Thille & Zimmaro, 2017) or successive summative scores (Ellis, 2013; Thille & Zimmaro, 2017). In itself, this data is not specific enough to determine follow-up actions (Thille & Zimmaro, 2017). Mere activity tracking insufficiently informs specific teacher decisions (Kong, et al., 2014). In combination with specific activity results however, registrations of time spent and the number of attempts become informative of student performance (Tempelaar, et al., 2015), and consequently relevant to embedded assessment.

Data versus Information

Although data and information are sometimes used as synonyms (e.g. Jasanoff, 2017), others distinguish raw, unprocessed data from *information* that results from data processing such as categorizing, summarizing and relating data to the predetermined goal (e.g. Davenport & Prusak, 1998; Jones, 1998 in Johnston & Kristovich, 2000; Marsh, Bertrand, & Huguet, 2015; Selwyn, Henderson, & Chao, 2015). Marsh et al. (2015) situate transformation of data to information in a *data use cycle* as a prerequisite step towards data use. Embedded assessment as conceptualized in this research presents the lecturer with information after processing the raw data, relating them to the learning objectives and related evaluation criteria included (Figure 1).

Actionable Knowledge as a Fundament for Action

Assessment information is fundamental to formative and summative decisions (Black & Wiliam, 2018). For this purpose, information is to be transformed into *actionable knowledge* that directs decisions and actions, by framing it in existing knowledge about learning and instruction (Marsh, et al., 2015; Wilson & Sloane, 2000). Traditionally, this transformation from information into actionable knowledge happens through interpretation by the lecturer. Over time, alternatives have emerged in the form of educational technology based on, for instance, artificial intelligence (Popenici & Kerr, 2017). Artificial tutors already can effectively support the acquisition of basic knowledge and skills. Learning activities of a higher order and with greater complexity cannot (yet) be guided by these tutors. In these more complex situations, interpretation and decision making by a human lecturer remains crucial (Luckin, 2018; Popenici & Kerr, 2017).

While interpreting and determining action, the lecturer or the application of artificial intelligence renders meaning to the information in terms of learning and instruction. Based on that, an adequate action is selected (Farrell & Marsh, 2016; Wilson & Sloane, 2000). Whereas artificial intelligence uses a set of pre-programmed rules (possibly extended with rules created by itself) (Luckin, 2018; Popenici & Kerr, 2017), a lecturer uses his or her personal frame of reference (Coburn & Turner, 2011). The current study's focus is on embedded assessment of complex skills in which information is interpreted and action determined by the lecturer (Figure 1).

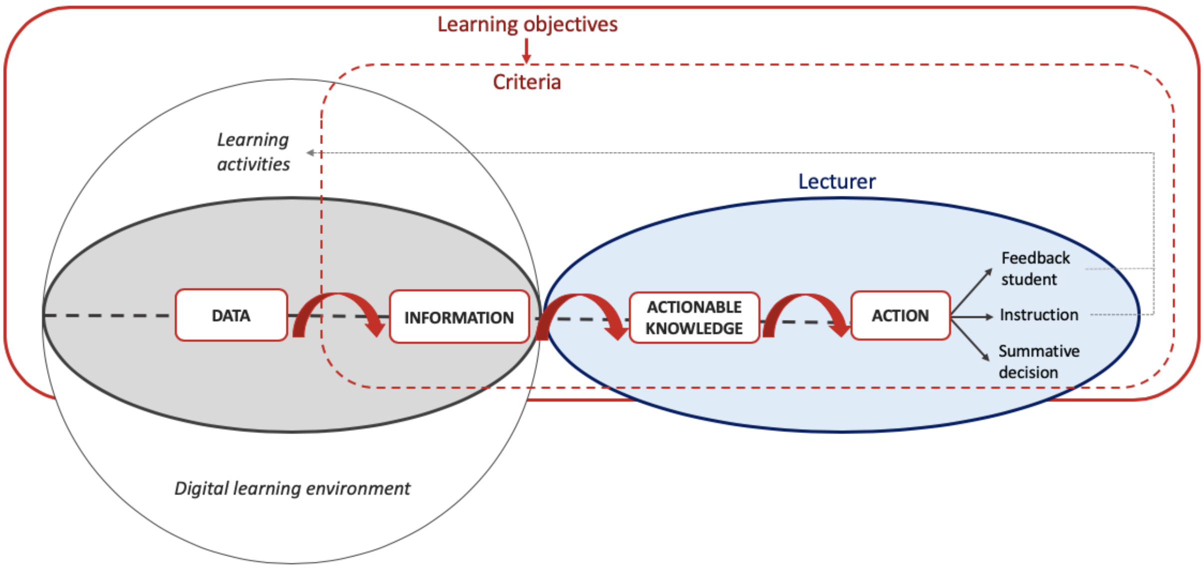

Figure 1. Embedded assessment: from data to information, to actionable knowledge and actions

Effective embedded assessment of complex skills facilitates the lecturers' transformation of delivered information into actionable knowledge by its concrete design. Moreover, the lecturers want to use this actionable knowledge to determine consequential instructional steps, provide feedback to the student and/or make a summative decision about the degree to which the student attained the predetermined learning objectives. Alignment of the provided information with lecturers' characteristics is important, as the next section further explains.

Aligning Embedded Assessment Design with Lecturers' Characteristics

Lecturer characteristics influence whether lecturers perceive information as useful and how they make decisions through interpretation (Coburn & Turner, 2011; Flower, 1985; Zapata-Rivera & Katz, 2014). Coburn and Turner (2011) name lecturers' beliefs, knowledge and motivation as determinants. Similarly, Flower (1985) mentions attitude, knowledge and needs. In this study we'll use the denominators (a) *beliefs* (attitude), (b) *knowledge* and (c) *needs* (motivation). We briefly elucidate each.

Lecturers' beliefs.

Positive beliefs regarding a particular assessment, more specifically the conviction that the assessment is meaningful, coincides with using assessment information (Heritage, 2007; Jonson, Tompson, Guetterman, & Mitchell, 2017; Young & Kim, 2010). Likewise, beliefs about the usefulness of specific data - determined by perceived data validity and, more broadly, perceived quality of data collection - affects lecturers' paying attention to the resulting information (Farrel & Marsh, 2016). Finally, it has been pointed out that recent developments like learning analytics can cause lecturers to fear they might no longer play a crucial role in decision making (Thille & Zimmaro, 2017).

Lecturers' knowledge.

Besides lecturers' beliefs, lecturers' knowledge is influential as well (Coburn & Turner, 2011; Flower, 1985; Huguet, Marsh, & Farrell, 2014; Marsh, et al., 2015; Zapata-Rivera & Katz, 2014). Various authors distinguish different types of knowledge, required from lecturers:

- domain specific knowledge concerning *learning progressions*: subsequent steps to be taken in acquiring domain specific objectives (Harshman, 2015; Heritage, 2007);
- knowledge of appropriate instructional strategies (Coburn & Turner, 2011; Heritage, 2007);
- knowledge of students' competences acquired prior to course enrolment (Heritage, 2007);
- knowledge of (quality) characteristics of the (formative) assessment type(s) used (Heritage, 2007); and
- knowledge of data analysis (Coburn & Turner, 2011).

When designing an embedded assessment, it is important to understand the actual knowledge base of the lecturers and what complementary knowledge they need (Bolhuis, Schildkamp, & Voogt, 2016; Flower, 1985; Zapata-Rivera & Katz, 2014).

Lecturers' needs.

Finally, lecturers ask for information they consider relevant: information that provides an answer to what they want to know and accomplish (Flower, 1985; Zapata-Rivera & Katz, 2014). The *perceived* relevance is not necessarily consistent with the absolute relevance in terms of constructive alignment: the degree to which learning objectives, learning activities, and assessment are in line with one another

(Biggs & Tang, 2011; Buck, Ritter, Jensen, & Rose, 2010; Patton, 2008, in Zakocs, Hill, Brown, Wheaton, & Freire, 2015; Sharkey & Murnane, 2003). Effective embedded assessment provides information that - through solid design - aligns with instruction (learning activities) as well as learning objectives (Shute, et al., 2016) and, in addition, with the lecturers' specific information needs arising from the actions they want to take (Hopster-den Otter, Wools, Eggen, & Veldkamp, 2017).

A term that is used to describe the degree to which information provided corresponds with lecturers' information needs, is action orientation (Jonson, et al., 2017; Patton, 2008, in Zakocs, et al., 2015; Weiss & Bucuvalas, 1980). Actionable information connects with the lecturer's needs and knowledge (Patton, 2008, in Zakocs, et al., 2015).

Various authors describe actionable information criteria in general terms, such as informing about strengths and weaknesses of individual students (Jonson, et al., 2017), signalling what specific knowledge and skills the student should acquire next (Goodman & Hambleton, 2004), and offering specific suggestions how to adjust current educational practices (Hamre & Capella, 2015; Weiss & Bucuvalas, 1980). Based on a focus group with primary education teachers, Hopster-den Otter, et al. (2017) explicitly state that the lecturer's specific information needs depend on the action he or she wants to take. Participants in their study mention a variety of information needs (or combinations thereof) depending on five types of action they distinguish. For differentiation purposes, for instance, teachers would like to receive information that positions each student relative to the learning objectives and/or recommends subdivisions of students in different instruction groups. When preparing for a feedback dialogue, they indicate to need additional information about the individual student's solution strategy, strengths and weaknesses, information about domain specific learning progressions and information about student characteristics such as motivation.

In addition, information can only be actionable if it is provided timely: it has to be quickly available to enable lecturers to carry out necessary adaptations (Heritage, 2007; Patton, 2008, in Zakocs, et al., 2015; Popham, et al., 2014; Sharkey & Murnane, 2003). Data that are delivered with delay, are experienced as less useful (Farrel & Marsh, 2016; Young & Kim, 2010).

Based on the above description of how lecturers' information needs influence the translation of information into actionable knowledge and related actions, a number of general design principles for embedded assessment can be formulated:

- the provided information is relevant (i.e. related to) learning objectives (constructive alignment);
- the provided information aligns with the lecturer's specific information needs; and
- the provided information is available when the lecturer needs it.

Another important design principle is that information presentation fosters information comprehension and use (Card, 2009; Vieira, Parsons, & Byrd, 2018). However, information design is a field of expertise in its own, which stretches beyond the scope of the current study, which focus is on what information should be presented, rather than how it should be presented.

Research questions

This study explores lecturers' beliefs and needs regarding the use of embedded assessment in their own educational practice. Four questions have been leading this exploration:

1. What do lecturers think of the idea of embedded assessment?
2. What added values and what limitations and risks do they perceive with respect to embedded assessment in general?
3. What, if any, opportunities do they see for using embedded assessment in their own course?
4. Would they want to use embedded assessment?

The research focusses specifically on *online* higher education, as this educational context pre-eminently permits large scale data collection; a crucial prerequisite to embedded assessment (Ellis, 2013). An additional advantage is that lecturers in online education programs are familiar with educational technology. Because of this, the gap between the current situation and the potential embedded assessment scenario is smaller than in regular higher education institutions, meaning that participants can more easily relate to potential future scenarios from their actual professional experience.

Method

Design

This study is rooted in the tradition of design research:

> "The systematic analysis, design and evaluation of educational interventions with the dual aim of generating research-based solutions for complex problems in educational practice, and advancing our knowledge about the characteristics of these interventions and the processes of designing and developing them." (Plomp, 2013, p. 16)

Design-based research typically consists of iterative cycles of problem analysis, design, development, and evaluation, leading – through a process of cyclical refinement - to both the development of a specific educational solution and the formulation of design principles (Edelson, 2002; Plomp, 2013).

The research described constitutes part of the first phase of systematic analysis, exploring the problem in general (see Introduction) and carrying out an analysis of beliefs and needs regarding embedded assessment from the perspective of one prominent stakeholder/user category: the teachers. Based on this analysis we derive tentative design principles. These can later be refined as a result of the needs analysis of other relevant stakeholders, for instance the students. The resulting tentative design principles inform the subsequent steps of (prototype) design, development, implementation, and evaluation.

To carry out the first analysis of the beliefs and needs of lecturers, we use a qualitative exploratory research design. In individual open-ended interviews, we question the lecturers' attitude towards the idea of embedded assessment, the added values, limitations and risks they perceive, as well as whether (and how) they would want to use embedded assessment in their own educational practice. At the start of a design process, with limited information available, open ended interviews with (future) users with experience in the relevant discipline/ function are pre-eminently suitable to explore design requirements (Carrizo, Dieste & Juristo, 2014). Participants can express their individual perception uninfluenced by the opinion of others, have the opportunity to bring up all topics they think about related to the subject and can ask questions instantly (Creswell, 2014). In addition, a researcher with limited training and experience in eliciting design requirements and with emerging familiarity with the research domain – as is the case in thesis research – can adequately elicit requirements through interviews (Carrizo, et al., 2014).

At the start of the interview, we show participants an introductory video, aimed at informing the participants about the central subject and scope of the interview. Using a video recording guarantees that each participant receives exactly the same information.

Participants

Rather than approaching an a-select sample of lecturers of the Open University of the Netherlands (OUNL), it was decided to initially limit the sample to lecturers of the master of Educational Sciences (N=33). These participants are lecturer as well as educational expert. Their thorough educational knowledge base is expected to result in relative high-quality results. In a sense, one could speak of *critical case sampling*: if it turns out hard already for this group of lecturers to elaborate on the leading research questions, it may well be infeasible for lecturers in other fields (Creswell, 2014; Etikam, Musa, & Alkassim, 2016). If the chosen research design turns out fine with this group of lecturers, then future expansion to other lecturer groups can be considered to explore tendencies in beliefs and needs throughout the institution.

A total of nine lecturers (27%) agreed to be interviewed: five women and four men, of various ages (M = 44.67, SD = 12.18), and years of experience, ranging from less than 5 to over 20 years (Table 2).

Table 2

Participants' professional experience as a lecturer

Years of experience	Number of participants
≤ 5	1
]5-10]	4
]10-20]	2
> 20	2

Materials

The materials used during the interview are firstly, an interview protocol and secondly, an introductory video explaining the concept of embedded assessment.

The interview protocol (Appendix A) entails that the interviewer prepares for the interview by studying the current assessment practice of the course(s) the participant is involved in, so that in this respect no detailed explanations are required and the interview can focus entirely on the questions at hand. Considering the interview itself the protocol describes four phases, further elaborated under Procedure:

- Phase 1: Opening the interview and registration of the participant's background characteristics through structured questions.
- Phase 2: Explaining embedded assessment by means of the video and offering an opportunity to ask further questions.
- Phase 3: Investigating the lecturer's beliefs and needs through open questions, based on the research questions.
- Phase 4: Closing the interview.

The structured questions for participants' background characteristics during phase 1 are: age, years of experience lecturing, the courses the participant teaches in and if the participants in these courses has the role of examiner (the latter two merely to verify whether course information found online is still up to date).

The introductory 6-minute video as part of phase 2 explains the concept of embedded assessment: the purpose and characteristics of embedded assessment and embedded assessment data, illustrated with three examples of embedded assessment practices. The introductory video was pilot tested with two educational scientists affiliated to another higher education institution, without pre-existing knowledge of the concept of embedded assessment. Based on their experiences and feedback, final adaptations were made.

Further, five *open questions* based on the research questions guide phase 3 of the interview:

1) Having seen the video, what do you think of embedded assessment?

2) Do you see added value in this approach? In what sense?
3) Do you see any limitations or risks?
4) Do you perceive opportunities to use embedded assessment in your own course(s)?
5) Would you want to use embedded assessment in your own course(s)?

Related to each of these questions, the interview protocol contains suggestions for specific follow-up questions.

The protocol contains space for the researcher to take structured notes. Regarding named opportunities for embedded assessment, note taking space is structured as a table, to ensure coverage of all central aspects identified in the theoretical background: (a) data, (b) information, (c) actionable knowledge, and (d) desired use/actions (Appendix A).

A trial interview (pilot test) was held with one lecturer/educational expert affiliated to another institution for higher education. The interviewee's profile is comparable to that of the participants selected for our study (Turner, 2010). Based on the lecturer's feedback two minor changes to the interview protocol were made: (a) at the start of the interview, we explicitly inform the participant we do not expect answers pro or contra embedded assessment, both are fine, we are interested in the participant's opinion and ideas, and (b) we added some additional possible follow-up questions to stimulate participants to think about implications for lecturers as well as students.

Procedure

After the university's Research Ethics Committee (Open University, 2019b) approved of the research proposal, lecturers of the master of Educational Sciences were invited by e-mail to participate in our research. After one week, a reminder was sent.

Interviews were held either face to face (at the Open University of the Netherlands) or online through videoconferencing using Collaborate, according to the participant's preference. The option to use video-conferencing expands the available time slots at which interviews can be held, especially given the geographical distance between interviewer and interviewee, while preserving visual advantages analogous to face-to-face interviews (Sedgwick, & Spiers, 2009).

Interviews consisted of four phases, as previously mentioned describing the interview protocol:

- Phase 1: Opening the interview and registration of the participant's background characteristics through structured questions. The researcher introduces himself, informs about the research purpose, explains the interview procedure and asks for the participant's informed consent, including permission for an audio recording. Subsequently, the researcher asks for the participant's background characteristics.
- Phase 2: Explaining embedded assessment by means of the video and offering an opportunity to ask further questions. Participants watch the introductory video. Following

the video, further clarifications are given where needed. If the participant asks for clarification of aspects of the concept embedded assessment, the researcher is instructed to provide it. If the participant seems to ask about the added values, limitations, risks or opportunities, the researcher is instructed not to answer the question in order to avoid influencing the participant's view. Instead, the researcher will try to verify this perception - for instance through mirroring the question: do you perceive this as a limitation or risk of embedded assessment?

- Phase 3: Investigating the lecturer's beliefs and needs through open questions, based on the research questions.
- Phase 4: Closing the interview: asking the participant for final remarks or questions about the central topic, thanking the participant and offering to send the participant a copy of the research findings (the thesis).

Going through these four phases took 30 minutes to 1 hour.

Data analysis

Interviews were transcribed verbatim in order to limit data reduction as much as possible (McLellan, MacQueen, & Neidig, 2003). No punctuation was inserted, because this could influence interpretation (DiCicco-Bloom, & Crabtree, 2006). Nonverbal sounds were included between brackets. All interjections of interviewer and interviewee were systematically included, except for interviewer humming synchronous to the interviewee's speech. For the sake of readability, this was omitted. Information that might lead to identification of the participant (e.g. age, course titles, gender) was substituted by a more general phrase (McLellan, et al., 2003). An example of this is "(…) in the eh course *[Title of course]* there is a final learning task in which (…)". In the exceptional case a number of words could not be distinguished, *(number of words not distinguished)* was inserted. To finalize transcription, a two-pass-per-tape policy was adopted, i.e., the transcript was checked twice listening simultaneously to the audio recordings (McLellan, et al., 2003).

Next, transcripts were manually coded with *in vivo* codes using the participant's own wording. An Excel document was constructed, containing all in vivo codes with related verbatim passages and a reference to the specific transcript the passage stems from. In the same document, the researcher grouped similar in vivo codes under overarching labels. Subsequently, each record (row), consisting of the verbatim passage with in vivo code and overarching label, was assigned to a theme, i.e. one of the central concepts of the theoretical framework (Creswell, 2014). These themes were: learning objectives/ activities, data, information, actionable knowledge, feedback, instruction, summative decision. Records that did not subsume under one of the central concepts, were grouped into the theme *embedded assessment in general*. Further analysing the results for each of these themes, a subordinate data classification gradually emerged through a process of interpretation and sparring with the thesis supervisor. Finally, tentative design principles for embedded assessment were derived.

Results

This section describes results consecutively addressing the four research questions regarding: (a) general beliefs, (b) perceived added values, limitations, and risks, (c) opportunities to use embedded assessment within own course, and (d) the wish to do so. Added values, limitations and risks are described in relation to each of the central concepts depicted in Figure 1. Some descriptions are illustrated with literal quotations from the interview transcripts. The translation of these originally Dutch quotations was made by the author.

General Beliefs about Embedded Assessment

Answering the question what they think of the idea of embedded assessment, six participants spontaneously expressed appreciation for embedded assessment using qualifications such as beautiful (*n*=3), good (*n*=2) or interesting (*n*=1). The others did not express any general, overall qualification. Four lecturers signalled a close relationship with concepts like learning analytics, stealth assessment, formative assessment and evidence informed instruction.

All but two lecturers explicitly stated that some learning objectives and activities are more suitable for embedded assessment than others. Six lecturers expressed thoughts along the line that unambiguous student actions that follow a fixed pattern are easier to model and hence, more suitable for embedded assessment, than complex actions, like context specific designs or creations. There was less agreement when it comes to learning objectives related to knowledge versus skills: two lecturers considered embedded assessment more suitable for assessing knowledge than skills, two other lecturers expressed exactly the opposite belief.

All participants spontaneously mentioned they consider embedded assessment as suitable for formative assessment purposes. Seven lecturers said embedded assessment might also be used for summative purposes. However, with respect to summative use, various conditions were explicitly formulated:

- the final decision must be made by a lecturer, as opposed to being based solely on artificial intelligence (*n*=2),
- embedded assessment should never be implemented for summative purposes only (*n*=2): *"I am not that much of an opponent to summative assessment but not that big of an advocate either unless it is a combination of formative and summative assessment" (Lecturer A).*
- embedded assessment must be based on a combination of different, relevant data (*n*=1):

"You probably will have to collect a variety of data, based on which you are better able to make such a decision" (Lecturer B).

- embedded assessment information must be complemented with information from other assessment types (*n*=1): *"If it is for summative assessment purposes I think you should consider very well what component it will be of your overall assessment I do not know at the moment I would not say that it should be based solely on it" (Lecturer C).*

Four lecturers questioned the feasibility of designing concrete embedded assessment practices. They especially considered the aspect that they are meant to be experienced by students as natural and less prominent (*n*=2).

Perceived Added Values, Limitations and Risks

Participants mentioned added values and/or limitations and risks related to data, information, actionable knowledge, and different types of actions, as well as on a more general level. We describe each of these in succession.

Added values, limitations and risks regarding data.

Table 3 summarizes the perceived added values, limitations and risks lecturers mentioned related to the type of data and/or the way data are collected. Perceived added values related to data richness and an expected positive effect of this particular type of data collection on student experience and study behaviour. Lecturers appreciated the synchronous collection of ubiquitous and abundant data as part of (embedded in) the learning process, including data about student actions in learning situations that cannot be directly observed by the lecturer. This embedded data collection was expected to positively affect student experience and study behaviour: it might reduce test anxiety and prevent students from cramming.

When discussing limitations and risks, limited data availability due to technological restrictions or student reluctance in using required tools is addressed as a point of concern. Current data collected in the existing digital learning environment was perceived as limited, although the participant signalling this aspect indicated not to know the learning environment's future potential. Also, using tools aside from those integrated in the digital learning environment might be experienced as bothersome by students.

Other risks/limitations named in relation to data (-collection) can be labelled as possible threats to assessment validity. Firstly, the quality of student modelling is influential: accurately determining what combination of data is most suitable in order to make valid, holistic inferences about a competence is challenging. A student model is always an approximation: *"it would presume the embedded assessment would have an ideal student model and that does not exist so you cannot model the student for 100%" (Lecturer D)*. Other anticipated validity threats are possibly distorted student behaviour during the assessment due to awareness of ongoing assessment or of the assessment model and a presumed augmented risk of fraud because of a lowered fraud threshold in case of data collection through technology.

In addition, lecturers discussed data related limitations or risks concerning reliability, transparency, privacy, ethics and negative student experience. With regard to reliability, the importance of data triangulation was underlined: *"I could presume that one also triangulates data during embedded assessment, that one also includes in one way or another the different actors that are involved in the entire process" (Lecturer B).* Also, a majority of lecturers perceived threats to transparency: the ubiquitous data collection might lead to lack of transparency towards the students if no explicit attention is paid to adequate communication about what data are collected, what criteria are used and what actions are based on the embedded assessment information. Questions regarding privacy and/or ethics were regularly raised. Finally, continuous data collection was signalled to possibly increase student stress or feeling of unsafety.

Table 3

Perceived added values and limitations/risks regarding embedded assessment data

Embedded assessment data			
Added value	*n*	Limitation /risk	*n*
Data richness		**Data availability**	
Data collection synchronous to / part of the learning process (embedded)	5	Technological data collection abilities of the (current) digital learning environment	1
Ubiquitous / including situations that cannot be directly observed	2	Student reluctance to use tools aside from those integrated in the digital learning environment	1
Abundant	1		
Unnoticed data collection	2	**Validity**	
Student experience and study behaviour		Quality of modelling	5
Reduces test anxiety	1	Distorted student behaviour due to awareness	3
Prevents cramming	1	Lowered threshold to fraud because of data collection by technology	1
		Reliability	
		Importance of data triangulation	1
		Transparency	
		Lack of transparency due to ubiquitous data collection	1
		Importance of transparently informing students	5
		Privacy	4
		Ethics, e.g. possibility to opt out	2
		Student experience	
		Increased student stress or feeling of unsafety due to continuous data collection	3

Note. Labels in bold are overarching labels assigned by the researcher.

Added values, limitations and risks regarding information.

Perceived added values of embedded assessment information relate to assessment validity (information is detailed and about actual behaviour), and to the fact that information results from automated data processing. However, lecturers also pointed out various limitations/risks related to embedded assessment information (Table 4).

Most frequently, limitations/risks described potential threats to assessment validity. Firstly, students' thoughts were said to remain a black box due to the fact that embedded assessment information is limited by the collection of data about traceable behaviour. These unexposed thoughts might however be relevant for competence assessment. Secondly, lecturers expressed concerns regarding the quality of the information derivation process: weighing and integrating different data and taking into account the variety of possible pathways and approaches in performing complex learning activities is perceived as challenging. Another element mentioned, is the quality of inferences enabled by technology: *"there is the limitation of what embedded assessment can do until it includes technology that one way or another would clarify the semantics of the relationship" (Lecturer D).*

Other limitations or risks concerned reliability or effectivity. A risk related to embedded assessment reliability was perceived in the possibly limited accuracy of technology recognizing and interpreting certain indicators: *"how good is software recognizing the different indicators that are considered important (…) software always is suboptimal" (Lecturer B).* Finally, determining the adequate degree of information specificity is a challenge that poses a possible threat to embedded assessments effectivity. Especially for information to be useful to substantiate a feedback dialogue, the desired degree of specificity depends on the student competence level: *"as students become a little more expert (...) you naturally hope you can communicate on a more abstract level and then I would hope embedded assessment can align with this so it provides adequate information" (Lecturer A).*

Table 4

Perceived added values and limitations/risks regarding embedded assessment information

Embedded assessment information			
Added value	*n*	Limitation or risk	*n*
Validity		**Validity**	
About actual behaviour	3	Students' thoughts remain a black box	3
Detailed	1	Quality of information derivation process	5
Automated data processing	1	**Reliability**	
		Of information derivation by technology	1
		Effectivity	
		Determining the adequate degree of specificity of provided information	1

Note. Labels in bold are overarching labels assigned by the researcher.

Added values, limitations and risks regarding actionable knowledge.

A frequently named added value of actionable knowledge gained through embedded assessment relates to its substance, i.e. the fact that it explicitly includes knowledge about the student's individual learning process as a basis for action as opposed to knowledge about student results only (Table 5). Besides, embedded assessment is expected to enhance objectivity of lecturer conclusions, as lecturers will base their judgements on the same provided information. Also, because of frequent and timely information provision, actionable knowledge is up-to-date. Finally, the cyclical nature of embedded assessment - data, information, actionable knowledge, actions, additional data, information… and so on - *"forces you to adopt a cyclical approach of learning, causing literally more informed future steps" (Lecturer E).*

Table 5

Perceived added values and limitations/risks regarding embedded assessment actionable knowledge

Embedded assessment actionable knowledge			
Added value	*n*	Limitation or risk	*n*
Substance		**Validity: required lecturer expertise**	
Knowledge about the student's individual learning process as a basis for action	6	Overwhelmed by extensive input, leading to loss of overview	2
Objectivity		Weighing and integrating information into actionable knowledge	1
Increased objectivity of lecturer conclusions	2	Thorough knowledge of learning progressions	1
Up-to-date	1	Blindly trusting delivered information	1
Cyclical	1	Interpreting assessment information without considering necessary context information	1

Note. Labels in bold are overarching labels assigned by the researcher.

Limitations and risks regarding actionable knowledge all concern the level of expertise required from lecturers to derive valid conclusions. Lecturers might be overwhelmed by the extensive amount of provided information, leading to loss of overview. Furthermore, weighing and interpreting various information into valid actionable knowledge is challenging. Adequate interpretation requires a thorough knowledge of learning progressions. Finally, there is the risk that a lecturer blindly trusts provided information and/or interprets it without considering necessary context information:

> *"it is different when the student for example has a husband who is seriously ill or something like that (…) but it can also be that someone just has a lot of trouble with the course level and that it actually is a kind of struggle to complete the course those are very different situations but with the same final result (…) there is the danger of concluding bluntly something that actually is not correct because you actually do not have certain background information that is necessary to take into account" (Lecturer C).*

Added values, limitations and risks regarding actions.

Perceived added values lecturers mentioned in relation to actions based on actionable knowledge, can be grouped according to the three types of action distinguished in the introduction: (a) providing the student with feedback, (b) adapting instruction and (c) summative decision making (Table 6). Interestingly, specific limitations and risks only concerned the action of providing feedback. However, with respect to the action of summative decision making this might be due to the fact that several lecturers already had expressed their thoughts/concerns regarding the use of embedded assessments for summative purposes, earlier in the interview (see General beliefs about embedded assessment).

Lecturers mentioned three added values which all appear to relate to feedback effectivity. Firstly, feedback dialogue can be facilitated as it is based on concrete student behaviour in relation to relevant assessment criteria. Secondly, the concrete behavioural mirror might serve to stimulate reflection. Finally, it was suggested that this type of embedded assessment feedback may help to reduce the Dunning-Kruger effect - which implies that poor performers overestimate the quality of their own performance (Kruger & Dunning, 1999) -: *"I think you could better counter the Dunning-Kruger effect (...) we know from the literature that one needs objective cues to be able to judge well how one is doing" (Lecturer A).*

In addition, embedded assessment feedback was judged as being timelier and more frequently available for students. Also, the majority of lecturers thought it interesting that embedded assessment feedback might (partly) be presented automatically and directly to the student. This would increase the availability of feedback even more without requiring additional lecturer effort. Finally, the experience of receiving personal feedback might foster academic integration.

In contrast, lecturers also mentioned limitations or risks related to feedback effectivity. In case of automated feedback presentation, feedback was characterized as canned, which might make it challenging to provide students with personal, context-specific feedback that facilitates learning. Moreover, student acceptance of automated feedback might be low. These suggested elements might all reduce the perceived value of automated feedback on personal student level. Besides, participants pointed at competences required on the part of students regarding both feedback interpretation and reasoning about their own learning process. Finally, the importance of lecturer expertise regarding feedback dialogue was considered to counterbalance the added value of feedback dialogue facilitation.

Regarding the second action detailed in the conceptual model, adjusting instruction, lecturers expected they can be more responsive using embedded assessment, and can be more agile - *"quick" (Lecturer A), "regular" (Lecturer F)* - in doing so. They also mentioned facilitation of instructional differentiation for subgroups or individual students. Finally, aggregated embedded assessment information is also considered valuable in terms of retrospective course evaluation and improvement of course design.

Table 6

Perceived added values and limitations/risks regarding embedded assessment actions

Added value	*n*	Limitation or risk	*n*
Feedback			
Effectivity		**Effectivity**	
Facilitates feedback dialogue based on concrete student behaviour related to the criteria	2	Low perceived value of automated feedback on personal level	5
Provided concrete behavioural ‘mirror’ can stimulate student self-regulation	3	Student difficulty interpreting presented information	3
Reduced Dunning-Kruger effect	1	Student difficulty reasoning about own learning process	2
Availability for students		Required lecturer expertise regarding feedback dialogue	1
Increased timeliness	4		
Increased frequency	1		
Efficiency			
Allows automated feedback/information presentation to students	6		
Student experience			
More personal feedback experience fosters academic integration	1		
Instruction			
Responsiveness			
Facilitates instructional adaptation	7		
Facilitates agile action	5		
Differentiation			
Instructional differentiation for individual students and/or student groups	6		
Course evaluation and design			
Useful input to retrospectively evaluate and improve course design	2		
Summative decision			
Substantiated decision			
Based on concrete, fine grained evidence	1		
About the learning process	2		
Reliability			
Stimulates reliability of assessment as compared to data collection at one point in time	2		

Note. Labels in bold are overarching labels assigned by the researcher.

With respect to the third action, summative decision making, lecturers expected embedded assessments to substantiate decisions with concrete, fine grained evidence. This includes decisions about the learning process: *“capturing the entire process so to speak and not just judging at the end what the final situation upon completion of the assignment is” (Lecturer G).* Reliability of summative decisions was also expected to be enhanced as compared to decisions based on data collected at one point in time:

"this could provide a lot of points in time or even a continuous flow so you do not make a decision based on a bad day" (Lecturer A).

Added values, limitations and risks in general.

Some of the added values, limitations and risks participants addressed could not be assigned to a particular aspect of embedded assessment. Topics that emerged here are summarized in Table 7.

Lecturers expected the initial development of embedded assessment to temporarily increase lecturers' work load. Once implemented, six lecturers expected embedded assessment to reduce work load /increase efficiency as a (combined) effect of automated data analysis, provision of direct, automated student feedback, and/or anticipation of potential future problems. In contrast, two lecturers thought embedded assessment implementation might increase work load due to massive information processing and/or by providing student feedback if this is not automated.

Table 7

Perceived added values and limitations/risks regarding embedded assessment in general

Added value	*n*	Limitation or risk	*n*
Reduced lecturers' work load		**Increased lecturers' work load**	
Due to automated data analysis	2	During embedded assessment development	2
Due to direct and automated student feedback	2	Due to massive information processing	1
Due to anticipation of potential future problems	2	Due to providing student feedback if this is not automated	1
Effectivity of education		**Development costs**	3
Allows lecturer to focus in depth on specific substantive aspects of learning	2	**Technological resources**	
Creates opportunities to rethink lecturers' assignment	2	System capacity	2
Might give direction to curricular development through aggregated information	1	General, not further specified	2
University reputation	1		

Note. Labels in bold are overarching labels assigned by the researcher.

Various lecturers mentioned added value seemingly related to increased effectivity of education on a more general level. In case, for instance, feedback on writing skills is provided automatically, the use of embedded assessment could allow lecturers to focus more in depth on specific substantive aspects of learning. Another opportunity created by embedded assessment is the use of embedded assessment data for rethinking lecturers' individual assignments:

> *"Maybe you could use a different model of lecturer allocation in which a lecturer does not guide the student during the entire process but in which you could for example say oh this lecturer is really good at this part so let us allocate him to this, without a problem because he can also look back as now processes become explicit that would otherwise actually not be visible" (Lecturer C).*

Considering the effectivity of the educational program, aggregating embedded assessment information from multiple courses might give direction to curricular development:

> *"You could of course compare groups or different years - I don't know how long this kind of data is kept – but you can compare and see how the curriculum develops or should develop and identify elements that need adaptation" (Lecturer G).*

Apart from these considerations related to effectivity and efficiency (work-load), there is the argument of reputation: the implementation of embedded assessment could contribute to university profiling.

General limitations mentioned by lecturers (apart from the possibly increased workload already addressed above) were related to resources: the costly development of embedded assessment as well as the demands on technological resources, i.e. system capacity. Whether the benefits of embedded assessment outweigh those costs was underlined as an important question to address prior to development.

Perceived Opportunities for Embedded Assessment in Own Course

All participants discussed one or more opportunities for embedded assessment in their own course(s), adding up to a total of 26 opportunities, most within the scope of current learning objectives, others in addition to current learning objectives. Appendix B provides an elaborate overview of each opportunity in terms of the learning activities, data, information, actionable knowledge, and actions involved. These opportunities can be grouped according to their distinct focus. Some focus on specific learning outcomes that are reflected in a concrete product the student creates, others on self-regulation or collaborative learning. Discussing these perceived opportunities, some lecturers also expressed general information preferences.

Perceived opportunities with regard to learning outcomes reflected in a concrete product. With regard to learning outcomes that are reflected in a concrete product the student creates, seven suggested opportunities focus on academic writing. Four lecturers would like to use embedded assessment information to derive conclusions about the current quality of the student's writing and to gain insight in the student's writing process leading to this result. One lecturer would want to know to what extent the student correctly connects different educational concepts and substantiates claims with literature and with concrete educational practices when writing critically about an educational topic. This in order to provide the student with automated feedback. Finally, two lecturers would like to decide if the student's work is eligible for summative assessment based on formal requirements. Data consists of written text in assignments. By comparing this data to quality indicators such as coherence, consistency, sentence structure and length, variance in writing style and correct spelling, information can be provided on different elements of writing quality. In addition, written text data could be compared to formal requirements such as use of the provided template, the expected number of words and compliance with APA directives.

Five perceived opportunities for embedded assessment contribute to actionable knowledge regarding learning activities involving an educational design task. Lecturers would like to make derivations about:

- how the student operationalizes design principles or specific logical design rules;
- to what extent the student concretizes the design in specifically described learning activities;
- how the student processes source material; and/or
- what line of reasoning the student follows concerning educational design.

Students would be asked to perform delimited steps of the design process in an online environment or to write a design report. During this design task, students' online actions are tracked and/or written text data are collected. Comparison of these data to design principles, provided study materials, common mistakes and/or other quality indicators, e.g. the degree of concretisation, results in specific embedded assessment information. Only formative action, more specifically providing (automated) feedback, was mentioned in the context of these opportunities related to educational design.

With regard to information skills, studying literature was mentioned twice as an interesting embedded assessment opportunity: one lecturer wanted to know how the student searches for literature, another lecturer wanted to know which (additional) materials the student studied during the course. References in the student's written assignments are the only data that were specified to this end. This and other data could be processed into an overview of studied material.

One lecturer suggested embedded assessment might be interesting with regard to observation skills in learning activities in which students observe a video recorded instructional situation. The processing of eye tracking data could provide information on the elements the student observed as compared to the elements marked as important by the lecturer. This information was thought interesting for providing automated feedback.

Discussing specific opportunities, five lecturers elaborated on opportunities to keep track of the student's progress as reflected in consecutive products in courses with related learning objectives. Therefore, they stated, embedded assessment should be implemented in the context of these multiple courses. Writing skills are mentioned three times as a focus, presentation skills and research skills each are mentioned once. Specific data mentioned with this regard are multiple recorded presentations of the same student throughout the educational program. By analysing and comparing these, the student's progress regarding presentation skills can be revealed. This can be used for direct feedback. Likewise, information about multiple written assignments across time can be displayed in a chronological overview. Another participant suggested to incorporate aggregated information e.g., in a student file or portfolio, that transfers with the student from one course to the next. This might be an overarching rubric or a graph visualizing student's growth.

Perceived opportunities with regard to self-regulation and collaborative learning. Three perceived opportunities are related to the student's self-regulation skills. Actionable knowledge lecturers would like to gain here concerns:

- the quality of the student's self-assessment;
- the amount of feedback the student needed to obtain the learning objectives;
- how the student acted upon received feedback; and
- how the student self-regulates in general.

Data mentioned by lecturers to be used in support of these opportunities include feedback the student received, student's self-evaluation of performance/result and the lecturer's evaluation of the student's performance/result. The latter two data are suggested to be input for a discrepancy analysis. Other information perceived as useful is the amount of feedback the student received, the sequence in which different elements of received feedback are addressed and a comparison of changes in the student's behaviour or work with received feedback. Two lecturers perceived opportunities for summative decision making, the other lecturer wanted to stimulate student reflection by taking formative action.

Collaborative learning activities were the focus of five mentioned opportunities. Lecturers expressed a desire for actionable knowledge about:

- the level/profoundness of joint knowledge development;
- what personal knowledge the student acquires during group work;
- how the student cooperates with peers;
- how cooperation during the learning process reflects in the final assignment; and/or
- the quality of the peer feedback provided by the student.

Data suggested were video recordings of student conversations, written contributions to the discussion forum or peer ratings of the student's feedback. Automated data processing was suggested to lead to information on the type and frequency of student interactions, average peer evaluation scores and/or the extent to which peer feedback or characteristics of studied peer assignments are reflected in the student's final assignment. In most cases (four out of five) the lecturer would want to take formative action, e.g. adaptively and automatically presenting specific questions or triggers during student conversations, referring the student to a specific peer or worked example or organizing an additional student-lecturer conversation. For one opportunity focussing on the quality of the student's peer feedback and one considering collaboration during group work, the embedded assessment was considered interesting for summative decision making.

General information preferences expressed while discussing perceived opportunities.

Discussing perceived opportunities for embedded assessment, some lecturers more generally described the information they would consider useful to receive. They elaborated on suggestiveness, aggregation level and/or timing of the provided information.

Regarding the suggestiveness, one lecturer asked for provided information to include suggestions for next instructional actions. Including a non-binding advice or a suggestion for summative decision making, was put forward twice. Apart from substantive information about individual students, three lecturers asked for information to facilitate differentiation, e.g. the distinct patterns or subgroups of students. Four lecturers wanted to receive aggregated information about the entire group, including the relative position of individual students, outliers or a comparison of group information from different cohorts.

With regard to the timing of information provision, two lecturers expressed a preference for information to be provided on demand ("pull" rather than "push"). One of them would complement this with set predefined moments during the course at which certain information is 'pushed' for example when a student hands in an assignment.

Desire to Use Embedded Assessment in Own Course(s)

All participants confirmed they would want to use embedded assessment during their course(s). Discussing this further, the majority of participants explicitly underlined a condition: they asked for evidence for the quality (validity and/or reliability) of the specific embedded assessment. Two lecturers mentioned they would want to explore student opinions and attitudes, e.g. through pilot testing, before deciding whether or not to use embedded assessment in their course.

Conclusion and discussion

This thesis explores embedded assessment from the perspective of lecturers of a higher online program in Educational Sciences. The investigation of their views and needs was preceded and guided by development of an embedded assessment framework based on the literature. This chapter summarizes the results for the main research questions and briefly discusses them in light of related research, before reflecting on the implications in terms of tentative design principles.

The first research question addresses general views of embedded assessment. The interviewed lecturers tend positively towards the concept of embedded assessment, especially in function of formative assessment purposes. With regard to summative assessment purposes and embedded assessment of learning activities with high cognitive complexity, they are more reserved. Interestingly, lecturers' reservation towards embedded assessment seems perpendicular to the view of Redecker and Johannessen (2013), who explicitly emphasize the necessity of an evolution towards embedded assessment in order to assess and facilitate the development of highly complex skills. Likewise, Schute et al. (2016) situate embedded assessment as integrated technology-enhanced assessment based on natural digital activity while tackling complex tasks.

The second question explores perceived added values and limitations or risks. Perceived added values mentioned by the lecturers fully cover the definition and purpose of embedded assessment as elaborated in the introduction of this thesis: embedded assessment is the *continuous*, nearly *unnoticed* collection of data about *observable aspects* of the student's knowledge, skills and attitudes during multiple learning activities, leading to a *substantiated* map of the student's learning *process and current competence level*. This facilitates *targeted feedback*, *agile adjustment of instruction* and *summative judgement* (Farrell & Rushby, 2016; Johnson-Glenberg, 2010; Redecker & Johannessen, 2013; Shute & Kim, 2013; Shute, et al., 2016; Shute, et al., 2009; Wilson & Sloane, 2000). In addition to these (italicized) aspects of this definition, lecturers value the automated data analysis; actionable knowledge being up-to-date, more objective and cyclical; and the facilitation of differentiation. Also, they appreciate the possibility of providing students with automated – and therefor more frequent - feedback and they expect summative decisions to be more reliable. These views are consistent with Drachsler and Greller's (2012) findings in a survey of international educational professionals - mainly from tertiary education –, who expected learning analytics to increase the speed of information about learning progress and to enhance objectivity of assessment.

When talking about embedded assessment limitations and risks, lecturers express multiple concerns regarding embedded assessment validity. Lecturers are well aware of the difficulty of modelling student progress, especially considering the extensive possibilities of concrete evidence that can be expected in diverse learning activities and learning paths. Shute, et al. (2016) consider this an important current limitation to embedded assessment to be addressed in future research. Likewise, Shavelson (2009) suggests that thorough research on learning progressions should precede incorporation of these learning progressions in assessment design. This because learning progressions depend on the instruction and the specific subsequent problems a student is confronted with in combination with the student's pre-existing related knowledge. A particular additional issue related to modelling is raised by Kane and Tannenbaum (2016), who criticize the relevance of performance information collected early during the learning process for certification purposes at course completion time.

Lecturers also mentioned a remaining black box containing students' thoughts as a perceived limitation to embedded assessment validity. However, discussing assessment designs, authors point out the possibility of incorporating think out loud data (Thomas, Saroyan, & Dauphinee, 2011) or written data in a thinking journal (Blakey, & Spence, 1990) to complement directly observable behaviour. Such data might also be incorporated in embedded assessment, however, this approach was not mentioned/envisioned during the interviews.

Related to embedded assessment validity, lastly, lecturers stress the required levels of lecturer expertise in deriving actionable knowledge. Similarly, a review of empirical evidence found possible misinterpretation of information by the lecturer to be a weakness of learning analytics and educational data mining (Papamitsiou & Economides, 2014). Some lecturers spontaneously suggest lecturer support in interpreting information, as does Schouten (2017) in relation to data use more generally by teachers in higher education.

In addition to the added values and limitations stated above, lecturers anticipate a positive impact of embedded assessment on curriculum development and the university's reputation. Discussing the student perspective and lecturer work load however, positive as well as negative effects are expected.

With regard to the student perspective, embedded assessment literature describes similar positive expected effects as those suggested by the lecturers. Shute and Kim (2013) anticipate less test anxiety if the assessment is embedded in highly immersive learning activities. In addition, Shute, et al. (2016) suggest embedded assessment might diminish cramming. On the other hand, concerns lecturers raise from the student perspective are also addressed by various authors discussing extensive data use in the field of education: transparency of data collection and data processing, students' privacy and consent, clarity of the information provided, possible stress or feelings of unsafety due to continuous data collection, and student difficulty reasoning about assessment information (Drachsler & Greller, 2012; Kong, et al., 2014; Luckin, 2018; Muravyeva, Janssen, Dirkx & Specht, 2019; Papamitsiou & Economides, 2014; Shute, et al., 2016; Spector, et al., 2016; Wang, 2016).

Lecturers' views on the possible impact of embedded assessment on lecturer work load are ambiguous. The development of embedded assessment is expected to temporarily increase work load, but the impact of embedded assessment once it is implemented is perceived differently: while the majority of lecturers expects a work load reduction, others anticipate work load to increase. Shute and Kim (2013) expect lecturer work load to decrease, allowing more time for formative lecturer action. While these authors generally attribute work load reduction to reduced time spent on students' assignments, lecturers suggest varied, more specific causes, i.e. automated data analysis, automated student feedback, and facilitated anticipation of potential future problems. Like Shute and Kim, some lecturers underline this would allow them to focus more in depth on specific aspects of student learning.

The third question in our study focusses on what, if any, opportunities lecturers see for using embedded assessment in their own teaching. All lecturers perceive opportunities for the use of embedded assessment in the context of their courses, be it – in some cases - after adapting the current learning objectives. Nearly half of the lecturers spontaneously put forward to implement embedded assessment throughout the curriculum in multiple courses addressing related learning objectives. Specific learning outcomes related to products the student creates as well as student self-regulation and cooperative learning are focusses of interest. A majority of lecturers mentions writing skills, the operationalization and concretization of design principles and/or cooperation with peers. Other suggestions include tracking the student's personal learning process and use of study materials. Remarkably, this list of perceived opportunities includes highly complex skills, which seems consistent with literature, but inconsistent with lecturers' general reservation regarding embedded assessment of this type.

In line with the expressed general beliefs about embedded assessment, perceived opportunities most frequently involve formative assessment purposes and to this end lecturers envision a rich variety of automated and non-automated feedback actions. A minority of opportunities, all related to self-regulation and collaborative learning, includes summative decision making. Additionally, deciding whether a student assignment is eligible for summative assessment is suggested as preferred action.

Depending on the desired actionable knowledge, lecturers require information about individual students, subgroups of students or aggregated information on the entire student group. A minority of lecturers asks for suggestions for next instructional steps or a non-binding advice for summative decision making to be included. This aligns with Popham, et al. (2014), who state that input of this type should never rule out that the lecturer decides.

In answer to the last question – whether the participants want to use embedded assessment - all expressed a willingness to use embedded assessment on the condition that its quality could be substantiated. Some participants would await further investigation, including elicitation of students' experience before or after a pilot test, before fully engaging.

Based on the interview results, some tentative design principles for embedded assessment can be formulated on various levels. Firstly, three general guidelines relate to the scope of embedded assessment:

- Consider the need to design cross curricular embedded assessment, spanning multiple courses, to allow longitudinal monitoring of student progress.
- Embedded assessment might include data providing insight in students' thought processes, e.g. think out loud data.
- Embedded assessment might also be usefully applied as a means to enhance efficiency of assessment procedures, e.g. to check the eligibility of an assignment for summative assessment (adherence to formal requirements).

Secondly, with respect to the implementation of embedded assessment in the specific context of the Educational Sciences program at the Open University, two specific foci appear particularly promising:

- Writing skills seem an interesting starting point, as they are the most mentioned application area for embedded assessment (and are mentioned also in relation to the previously named cross curricular scope).
- Besides, explore the possibilities for implementing course-specific embedded assessment of easy to model aspects of educational design.

Thirdly, lecturers mention three aspects of embedded assessment quality, which they need to be convinced of before considering adoption for their own practice:

- Design and implementation of embedded assessment should be evidence informed and guided by validated learning progression models.
- The technology used to collect and analyze the data must be proven to be reliable.
- Make sure it is transparent for the students what data are collected, what criteria are used and what actions might be expected based on what information.

Fourthly, from an information design / usability perspective, three principles can be formulated:

- Depending on specific contexts, information provisioning should allow lecturers to switch between views /aggregation levels (individual students, subgroups, total group, or a combination of these).
- Embedded assessment should provide information on a pull rather than push basis unless the lecturer specifies otherwise (e.g. a wish for information being pushed at pre-defined strategic moments).
- Check the students' competence in interpreting different types of assessment information and subsequently determine what specific information will be presented as direct automated feedback to the students, what feedback should always be part of a feedback dialogue and/or what additional student support interpreting assessment information might be needed.

Fifthly, three design principles consider further strategies to foster the successful implementation of embedded assessment:

- During a first implementation stage, use embedded assessment for formative assessment purposes only. Excluding summative assessment purposes in this first stage is aligned with lecturers' reservations towards summative assessment purposes and most of lecturers' perceived opportunities for embedded assessment in their own courses.
- Check if lecturers are well prepared for embedded assessment in terms of interpreting assessment information and provide customized lecturer support where appropriate.
- Explicitly ask student permission for the collection and processing of personal data where necessary.

Finally, in addition to the above tentative design principles, it goes without saying that embedded assessment design must balance available resources in terms of time, effort, system capacity, and availability of data.

The exploration of lecturers' beliefs and needs described in this study is limited to a small group of lecturers-educational experts. Above results are based on interviews with nine lecturers, who were interested to participate in research on this specific topic; some of them explicitly mentioned to be intrigued by the topic. The results indicate that overall, participants are critical but mainly positive towards embedded assessment. Considering the sampling method, a self-selection bias cannot be excluded (Robinson, 2014), and therefore results of this study cannot be generalized to lecturers in online higher education in general, nor to the entire team of lecturers of the master of Educational Sciences. However, they do reveal relevant aspects of lecturers' beliefs and needs with regard to embedded assessment in general and in the context of their own educational practice.

Future research could explore whether the beliefs and needs of lecturers at different educational programs and universities are similar to or differ from those expressed by this critical sample. With a view on this, some evaluative comments regarding the use of the framework for embedded assessment developed at the start of this study to introduce the concept of embedded assessment to the participants during the interviews might be considered.

During the interviews, some participants spontaneously, explicitly and correctly referred to specific elements of the framework, e.g. actionable knowledge. At other times, whether the participant perceived the distinction between data, information and actionable knowledge the same as conceptualized in the framework was unclear. This did not hinder data analysis, because analysis was based on consistently relating the actual content of participants' answers to the elements of the framework. However, a print of the framework being constantly available throughout the interview - contrary to restricting visualisation to the introductory video – might facilitate participants to adopt the different elements of the framework when thinking about and discussing embedded assessment.

Referring to the central position of the lecturer in the framework, one participant expressed uncertainty as to what degree the student perspective should be discussed during the interview. In addition, multiple participants asked whether the framework meant no information was provided directly to the student. In these cases, the interviewer explained that the student might also receive some information, but that this framework focused on the lecturer's use of embedded assessment. In future research, this can be clarified by adding this explanation to the introductory video.

The elements in the developed framework are closely related to the domain specific expertise of the participants. It is recommended to test the clarity of the framework to other potential user groups before deciding to use it as an introduction to the interview topic.

Finally, the principles formulated with a view on possible future embedded assessment design at this particular educational program, provide only a start. Further requirement elicitation is necessary, including the needs of other relevant stakeholders, certainly students. Based on this a number of user profiles representing the different stakeholder needs can be described. Subsequently, a multidisciplinary team consisting of educational designers and information technology experts can use these user profiles to determine a complete set of design requirements (Barré, Buisine & Aoussat, 2018).

References

Barré, J., Buisine, S., & Aoussat, A. (2018). Persona logical thinking: improving requirements elicitation for multidisciplinary teams. *CoDesign, 14*(3), 218-237. doi:10.1080/15710882.2017.1301959

Biggs, J. B., & Tang, C. S. (2011). Constructively aligned teaching and assessment: what is constructive alignment? in J. B. Biggs & C. S. Tang. *Teaching for quality learning at university: what the student does* (pp.97-100). Maidenhead, Engeland: McGraw-Hill Education.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. doi:10.1080/0969595980050102

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice,* 1-25. doi:10.1080/0969594X.2018.1441807

Blakey, E., & Spence, S. (1990). Developing metacognition. Retrieved from: https://www.ericdigests.org/pre-9218/developing.htm

Bolhuis, E., Schildkamp, K., & Voogt, J. (2016). Data-based decision making in teams: enablers and barriers. *Educational Research and Evaluation, 22*(3-4), 213-233. doi:10.1080/13803611.2016.1247728

Buck, S., Ritter, G. W., Jensen, N. C., & Rose, C. P. (2010). Teachers say the most interesting things: An alternative view of testing. *Phi Delta Kappan, 91*(6), 50-54. doi:10.1177/003172171009100613

Card, S. (2009). Information visualization. In A. Sears, & J. A. Jacko (Eds.), *Human-computer interaction: Design issues, solutions and applications* (pp. 509-543). Boca Raton, FL: CRC Press.

Carrizo, D., Dieste, O., & Juristo, N. (2014). Systematizing requirements elicitation technique selection. *Information and Software Technology, 56,* 644-669. doi:10.1016/j.infsof.2014.01.009

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement, 9,* 173-201. doi:10.1080/15366367.2011.626729

Creswell, J. W. (2014). Educational research: Planning, conducting and evaluating quantitative and qualitative research (4th ed.). Essex: Pearson Education Limited.

Cummings, R., Maddux, C. D., & Richmond, A. (2008). Curriculum-embedded performance assessment in higher education: Maximum efficiency and minimum disruption. *Assessment & Evaluation in Higher Education, 33*(6), 599-605. doi:10.1080/02602930701773067

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know.* Boston, MA: Harvard Business School Press.

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education, 40,* 314-321. doi:10.1111/j.1365-2929.2006.12418.x

Drachsler, H. & Greller, W. (2012). Confidence in learning analytics. *LAK12: 2nd International Conference on Learning Analytics & Knowledge.*

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences, 11*(1), 105-121. doi:10.1207/S15327809JLS1.101_4

Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology, 44*(4), 662-664. doi:10.1111/bjet.12028

Etikam, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1-4. doi:10.11648/j.ajtas.20160501.11

Farrell, C. C., & Marsh, J. A. (2016). Metrics matter: How properties and perceptions of data shape teachers' instructional responses. *Education Administration Quarterly, 52*(3), 423-462. doi:10.1177/0013161X16638429

Farrell, T., & Rushby, N. (2016). Assessment and learning technologies: An overview. *British Journal of Educational Technology, 47*(1), 106-120. doi:10.1111/bjet.12348

Flower, L. (1985). *Problem-solving strategies for writing (Second edition).* Orlando, FL: Harcourt Brace Jovanovich.

Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., & Yin, Y. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education, 21*(4), 360-389. doi:10.1080/08957340802347852

Gerretson, H., & Golson, E. (2004). Introducing and evaluating course-embedded assessment in general education. *Assessment Update, 16*(6), 4-6.

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education, 17*(2), 145-220. doi:10.1207/s15324818ame1702_3

Hamre, B. K., & Cappella, E. (2015). Measures of early adolescent development and school contexts: Narrowing the research to practice divide. *Journal of Early Adolescence, 35*(5-6), 586-596. doi:11.1177/0272431615578275

Harshman, J. T. (2015). *Characterizing high school chemistry teachers' use of formative assessment data to improve learning* (Doctoral dissertation). Retrieved from: https://etd/ohiolink.edu

Haug, B. S., & Ødegaard, M. (2015). Formative assessment and teachers'sensitivity to student responses. *International Journal of Science Education, 37*(4), 629-654. doi:10.1080/09500693.2014.1003262

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan,* 89(2), 140-145. doi:10.1177/003172170708900210

Hopster-den Otter, D., Wools, S., Eggen, T. J. H. M., & Veldkamp, B. P. (2017). Formative use of test results: A user's perspective. *Studies in Educational Evaluation, 52,* 12-23. doi:10.1016/j.stueduc.2016.11.002

Huguet, A., Marsh, J. A., & Farrell, C. C. (2014). Building teachers' data-use capacity: insights from strong and developing coaches. *Education Policy Analysis Archives, 22*(52), 1-26. doi:10.14507/epaa.v22n52.2014

Jasanoff, S. (2017). Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data & Society, 4*(2), 1)15. doi:10.1177/2053951717724477

Johnson-Glenberg, M. C. (2010). Embedded formative e-assessment: Who benefits, who falters. *Educational Media International, 47*(2), 153-171. doi:10.1080/09523987.2010.492681

Johnston, G. H., & Kristovich, S. A. R. (2000). Community college alchemists: Turning data into information. *New Directions for Community Colleges, 109,* 63-73. doi:10.1002/cc.10908

Jonson, J.L., Thompson, R. J., Guetterman, T. C., & Mitchell, N. (2017). The effect of informational characteristics and faculty knowledge and beliefs on the use of assessment. *Innovative Higher Education, 42,* 33-47. doi:10.1007/s10755-016-9366-7

Kane, M. T., & Tannenbaum, R. J. (2016). Some architecture for embedded-assessment systems. *Measurement: Interdisciplinary Research and Perspectives, 14*(4), 153-155. doi:10.1080/15366367.2016.1256158

Kerby, D., & Romine, J. (2009). Develop oral presentation skills through accounting curriculum design and course-embedded assessment. *Journal of Education for Business, 85*(3), 172-179. doi:10.1080/08832320903252389

Kong, S. C., Chan, T., Griffin, P., Hoppe, U., Huang, R., Kinshuk,…Yu, S. (2014). E-learning in school education in the coming 10 years for developing 21st century skills: Critical research issues and policy impications. *Educational Technology & Society, 17*(1), 70-78.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. doi:10.1037//0022-3514.77.6.1121

Libin, A., Lauderdale, M., Millo, Y., Shamloo, C., Spencer, R., Green, B., Donnellan, J., …, & Groah, S. (2010). Role-playing simulation as an educational tool for health care personnel: Developing an embedded assessment framework. *Cyberpsychology, Behavior, and Social Networking, 13*(2), 217-224. doi:10.1089/cyber.2009.0040

Luckin, R. (2018). Machine learning and human intelligence: The future of education for the 21st century. London: UCL Institute of Education Press.

Marsh, J. A., Bertrand, M., & Huguet, A. (2015). Using data to alter instructional practice: The mediating role of coaches and professional learning communities. *Teachers College Record, 117*(4), 1-40.

McLellan, E., MacQueen, K. M., & Neidig, J. L. (2003). Beyond the qualitative interview: Data preparation and transcription. *Field Methods, 15*(1), 63-84. doi:10.1177/1525822X02239573

Muravyeva, E., Janssen, J., Dirkx, K., & Specht, M. (2019). Students' attitudes towards personal data sharing in the context of e-assessment: informed consent or privacy paradox? In S. Draaijer, D. Joosten-ten Brinke, & E. Ras (Eds.), *Technology Enhanced Assessment* (pp. 16-26). Cham: Springer International Publishing. doi:10.1007/978-3-030-25264-9_2

Open Universiteit (2019a). *Wij zijn de Open Universiteit.* Retrieved from: https://www.ou.nl/over-ons

Open University (2019b). *Research Ethics Committee.* Retrieved from: https://www.ou.nl/research-ethics-committee-ceto

Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49-64.

Park, J. W., Seo, E. J., You, M., & Song, J. (2016). Development and application of course-embedded assessment system for program outcome evaluation in the Korean nursing education: A pilot study. *Nurse Education Today, 38,* 48-53. doi:10.1016/j.nedt.2015.12.018

Pike, G. R. (2014). Course embedded assessment designs: Lessons from the MSC. *Assessment Update, 26*(6), 7-10. doi:10.1002/au

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational Design Research* (pp. 10-51). Enschede: Netherlands institute for curriculum development (SLO). Retrieved from: http://downloads.slo.nl/Documenten/educational-design-research-part-a.pdf

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, *12*(22). doi:10.1186/s41039-017-0062-8

Popham, W. J., Berliner, D. C., Kingston, N. M., Fuhrman, S. H., Ladd, S. M., Charbonneau, J., & Chatterji, M. (2014). Can today's standardized achievement tests yield instructionally useful data: Challenges, promises and the state of the art. *Quality Assurance in Education, 22*(4), 300-316. doi:10.1108/QAE-07-2014-0033

Redecker, C., & Johannessen, O. (2013). Changing Assessment: Towards a new assessment paradigm using ICT. *European Journal of Education, 48*(1), 79-96. doi:10.1111/ejed.12018

Ridgway, J., & McCusker, S. (2008). Challenges for research in e-assessment. In F. Scheuermann & A. G. Pereira (Eds.), *Towards a research agenda on computer-based assessment.* Luxembourg: Office for Official Publications of the European Communities.

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology*, *11*(1), 25-41. doi:10.1080/14780887.2013.801543

Rose, L. T., & Fisher, K. W. (2011). Garbage in, garbage out: having useful data is everything. *Measurement: Interdisciplinary Research & Perspective*, *9*(4), 222-226. doi:10.1080/15366367.2011.632338

Schouten, G. (2017). On meeting students where they are: Teacher judgment and the use of data in higher education. *Theory and Research in Education, 15*(3), 321-338. doi:101177/1477878517734452

Sedgwick, M., & Spiers, J. (2009). The use of videoconferencing as a medium for the qualitative interview. *International Journal of Qualitative Methods, 8*(1), 1-11. doi:10.1177/160940690900800101

Selwyn, N., Henderson, M., & Chao, S.-H. (2015). Exploring the role of digital data in contemporary schools and schooling: '200.000 lines in an Excel spreadsheet'. *British Educational Research Journal*, *41*(5), 767-781. doi:10.1002/berj.3186

Sharkey, N. S., & Murnane, R. J. (2003). Learning from student assessment results. *Educational Leadership, 61*(3), 77-81.

Shavelson, R. J. (2009, June). Reflections on learning progressions. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A.,…Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education, 21*(4), 295-314. doi:10.1080/08957340802347647

Shute, V. J., & Kim, Y. J. (2013). Formative and stealth assessment. In J. M. Spector, D. M. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (4th ed.) (pp. 311-321). New York: Springer Science Business Media New York.

Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M.-W. (2016). Advances in the Science of Assessment. *Educational Assessment, 21*(1), 34–59. doi:10.1080/10627197.2015.1127752

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and Effects* (pp. 295-321). New York: Routledge.

Siemens, G., & Gasevic, D. (2012). Guest editorial: Learning and knowledge analytics. *Educational Technology & Society, 15*(3), 1-2.

Spector, J. M., Ifenthaler, D., Sampson, D., Yang, L., Mukama, E., Warusavitarana, A.,…Gibson, D. C. (2016). Technology enhanced formative assessment for 21st century learning. *Educational Technology & Society, 19*(3), 58-71.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167. doi:10.1016/j.chb.2014.05.038

Thille, C., & Zimmaro, D. (2017). Incorporating learning analytics in the classroom. *New Directions for Higher Education, 179*, 19-31. doi:10.1002/he

Thomas, A., Saroyan, A., & Dauphinee, W. D. (2011). Evidence-based practice: A review of theoretical assumptions and effectiveness of teaching and assessment interventions in health professions. *Advances in Health Science Education, 16*, 253-276. doi:10.1007/s10459-010-9251-6

Turner, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report, 15*(3), 754-760.

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: a systematic review and research agenda. *Computers & Education, 122*, 119-135. doi:10.1016/j.compedu.2018.03.18

Wang, Y. (2016). Big opportunities and big concerns of big data in education. *TechTrends, 60,* 381-384. doi:10.1007/s11528-016-0072-1

Webb, M., Gibson, D., & Forkosh-Baruch, A. (2013). Challenges in information technology supporting educational assessment. *Journal for information technology supporting educational assessment, 29*, 451-462. doi:10.1111/jcal.12033

Weiss, C. H., & Bucuvalas, M. J. (1980). *Social science research and decision-making.* New York: Columbia University Press.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*(2), 181-208. doi:10.1207/S15324818AME1302_4

Young, V. M., & Kim, D. H. (2010). Using assessment for instructional improvement: A literature review. *Education Policy Analysis Archives, 18*(19), 1-40. doi:10.14507/epaa.v18n19.2010

Zakocs, R., Hill, J. A., Brown, P., Wheaton, J., & Freire, K. E. (2015). The data-to-action framework: A rapid program improvement process. *Health Education & Behavior, 42*(3), 471-479. doi:10.1177/1090198115595010

Zapata-Rivera, J. D., & Katz, I. R. (2014). Keeping your audience in mind: applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice, 21*(4), 442-463. doi:10.1080/0960504X.2014.936357

Zenisky, A. L., & Hambleton, R. K. (2012). Developing test score reports that work: The process and best practices for effective communication. *Educational Measurement: Issues and Practice, 31*(2), 21-26. doi:10.1111/j.1745-3992.2012.00231.x

Appendix A

Interviewprotocol

Nodige materiaal:

- Dit protocol
- Het toestemmingsformulier ter ondertekening, in tweevoud
- Opnameapparaat, volledig opgeladen
- Laptop met videopresentatie, volledig opgeladen
- Balpen

(Vooraf te noteren door interviewer) Datum interview: ________________

1. Opening van het interview

Bedankt om in te gaan op de uitnodiging voor dit interview.

Mijn naam is Liesbet Snoeys. Ik ben student aan de master Onderwijswetenschappen hier bij Open Universiteit en werk daarnaast als onderwijsondersteuner aan de Karel de Grote-Hogeschool - of KdG - in Antwerpen. Dit interview kadert in mijn masterthesis aan de opleiding, die ik onder begeleiding van José Janssen uitwerk.

In het kader van mijn masterthesis onderzoek ik de opvattingen en behoeften rond embedded assessment – een begrip dat ik zo verder zal toelichten - bij de docenten (onderwijsexperts) aan de opleiding Onderwijswetenschappen.

*Het interview van vandaag duurt **maximaal 1 uur**. Eerst bekijken we een videopresentatie over embedded assessment, nadien leg ik je een aantal open vragen voor. Ik zou graag een **opname** willen maken van het interview, zodat ik me beter op het gesprek kan richten en niet tegelijkertijd aantekenen hoef te maken. Vind je dat goed? Ik gebruik daarvoor 2 apparaten, om een eventueel defect aan één van beide op te vangen. De verwerking van de data gebeurt **anoniem**. Tijdens het interview noteer ik ook af en toe **steekwoorden**. Dit doe ik om tijdens het luisteren overzicht te houden over wat je me al vertelde en op basis daarvan eventuele vervolgvragen te bepalen.*

Als ik het goed heb, ben je docent in de cursus(sen)

Neem je binnen deze cursus ook de rol op van examinator?

Mag ik je daarnaast als achtergrondinformatie ook je leeftijd vragen en het aantal jaren dat je ervaring hebt als docent?

2. Tonen van de videopresentatie

*De videopresentatie die ik je nu ga tonen bevat **toelichting** over het concept embedded assessment, geïllustreerd met een aantal **concrete voorbeelden**. De presentatie duurt **6 minuten**.*

Interviewer toont de videopresentatie. De presentatie bevat een beschrijving van het doel en de kenmerken van embedded assessment en embedded assessment data. Ter illustratie en concretisering bevat deze ook drie voorbeelden van mogelijke embedded assessment toepassingen.

*Is de idee van embedded assessment in deze presentatie **voldoende duidelijk** gemaakt? Heb je **nog vragen** naar aanleiding van deze presentatie?*

Instructies na het stellen van deze vraag: - Als de participant vragen stelt die een goed begrip van het concept 'embedded assessment' in de weg kunnen staan, verhelder dan wat voor de participant onduidelijk is. - Lijkt de participant op dit moment al naar meerwaarden, beperkingen of risico's te vragen, toets dan af of dit zo is: *"Zie je dat als een meerwaarde/beperking/risico van embedded assessment?"*

*De presentatie illustreerde een aantal typerende kenmerken van embedded assessment. Nu volgen een aantal vragen waarop **je vrijuit mag antwoorden**. Ik verwacht niet dat je pro of contra embedded assessment bent. Beide zijn goed. Ik ben **benieuwd naar jouw ideeën en mening.***

3. De opvattingen en de motivatie van de docent: hoofd- en vervolgvragen

3.1. Wat vind je van de idee van embedded assessment?

Mogelijke bijvragen:

- Wat maakt dat je dat vindt?
- Kan je dat toelichten?
- Zijn er andere zaken waar je aan denkt?

3.2. Zie je meerwaarde aan embedded assessment? In welke zin?

Mogelijke bijvragen:

- Kan je dat toelichten?
- Hoe zie je deze meerwaarde?
- Denk je nog aan andere manieren waarop embedded assessment een meerwaarde zou kunnen zijn?
    - Als je hierover denkt vanuit jouzelf als docent?
    - Als je hierover denkt vanuit de student?

3.3. Zie je beperkingen of risico's van embedded assessment?

Mogelijke bijvragen:

- Kan je dat toelichten?
- Hoe zie je deze beperking/ dit risico?
- Aan welke andere beperkingen of risico's denk je nog?
  - Als je hierover denkt vanuit jouzelf als docent?
  - Als je hierover denkt vanuit de student?

3.4. Zie je mogelijkheden om embedded assessment in jouw eigen cursus te gebruiken?

Mogelijkheden

Als de participant aangeeft mogelijkheden te zien in de eigen cursus, stel dan volgende vervolgvragen:

- Welke mogelijkheden zie je?
- Hoe zie je dit?

Noteer het antwoord van de participant in steekwoorden in onderstaande tabel en bepaal op basis daarvan specifieke vervolgvragen.

Geeft de participant aan geen mogelijkheden te zien, ga dan naar **‘geen mogelijkheden’**.

(Voorbeeld van concrete) data	**Informatie**	**Actionable knowledge**	**Gewenst gebruik/ acties**
Welke data zijn daarvoor nodig? Worden die op dit moment bijgehouden/ verzameld?	Welke informatie heb je daarvoor nodig?	Wat wil je uit het embedded assessment kunnen afleiden?	Welke acties zou je op basis hiervan willen nemen?

Zou je embedded assessment op deze manier willen gebruiken?

Geen mogelijkheden

Als de participant aangeeft *geen* mogelijkheden te zien in de eigen cursus, stel zo nodig dan volgende vraag:

- Kan je dit toelichten?

Vraag vervolgens (indien het antwoord van de participant dit nog niet bevatte):

- Zou je het wel willen? Waarom wel/niet?

4. Afronding

4.1. Zijn er aan het einde van dit interview nog zaken waar je op wil ingaan?

4.2. Heb je nog andere bedenkingen of vragen?

Tot slot

Van harte dank voor dit gesprek.

Na afronden van de scriptie zal ik je een exemplaar van het scriptieonderzoek toesturen.

Appendix B

Table B1

Perceived opportunities for embedded assessment in participants' courses

#	Learning activity	Data	Information	Actionable knowledge	Action
Opportunities with regard to specific learning objectives reflected in a product the student creates					
1	Writing assignment Multiple writing assignments throughout the educational program	Written text	Analysis of coherence, consistency and scientific writing	What is the quality of the student's writing? How does the student progress regarding writing skills?	Feedback
2	Writing an advisory report Multiple writing assignments throughout the educational program	Written text	Analysis based on writing quality indicators	What is the quality of the student's writing? How do the student's writing skills develop?	- (currently no learning objective)
3	Writing assignment	-	-	What does the student's writing process look like?	- (currently no learning objective)
4	Writing a structured design report Multiple writing assignments throughout the educational program	Written text	Analysis of Sentence structure Sentence length Variance in writing style Spelling mistakes Chronological overview of analysis at certain point in time	What is the quality of the student's writing? What does the student's writing process during one writing assignment look like?	Feedback (partially automated and partially direct)
5	-	Written text	Analysis of sentence structure	-	Automated direct feedback
6	Writing a critical appraisal or discussion about a specific educational topic	Written text	-	Does the student connect information or concepts? Does the student substantiate with literature and with personal professional practice?	Automated feedback (before submission)

7	Writing a design report	Written text	Comparison to APA directives Required global structure of the document Other formal rules	Does the student’s work meet the formal requirements to be eligible for assessment?	Automated direct feedback
8	Writing a design report	Written text	Check Use of the required template Completeness in terms of required components Required length Other formal rules	Does the student’s work meet the formal requirements to be eligible for assessment?	Start assessment or not
9	Designing instruction in an online learning environment	Live recording of actions taken and characteristics designed by the student	Student actions and design characteristics compared to common mistakes	How does the student operationalize design principles?	Automated direct feedback: signalling what went wrong and why it is wrong, providing hints or suggestions.
10	Developing a delimited step in the total design process	-	Design characteristics compared to specific logical rules	How does the student’s design reflect specific logical design rules?	Automated direct feedback
11	Writing a design report, describing the design of concrete learning activities based on a design model	Written text	Passages containing only global descriptions of the design model versus passages concretized with elements of the specific design	To what extent does the student translate global design principles into concrete learning activities in a specific context?	Feedback (before submission)
12	Writing a design report, describing the design of a specific educational practice based on source material about educational design	Written text	Semantic relation of the student’s text with the source material	How does the student process course material?	-
13	Writing a design report	Written text	-	What reasoning lines does the student follow?	-
14		-	-	How does the student search for literature?	- (currently no learning objective)
15	Spontaneously studying additional facultative material	References in written assignment	Overview of Studied additional	To what extent did de student search for	-

	and/or restudying material from previous courses		facultative material Restudied material from previous courses	and study additional facultative material? During this course, how did the student use material from previous courses?	
16	Observation of a video recorded instructional situation	Eye tracking data	Elements marked as important by the lecturer, that are observed by the student	-	Automated feedback: providing individualized specific additional observation instructions
17	Diverse presentation assignments throughout the educational program	Recorded presentations	Analysis of presentations	How does the student progress regarding presentation skills?	Direct feedback
18	Multiple assignments implying research skills	-	-	-	(currently no learning objective in own course)
Opportunities with regard to self-regulation and collaborative learning					
19	-	-	Amount of feedback the student received Speed of feedback processing Sequence of addressing the different elements of received feedback Changes in the student's behavior or work and their relation to received feedback	How does the student steer his/her behavior during the learning process? What amount of feedback does the student need to obtain the learning objectives and how did the student act upon it?	Summative decision
20	-	Feedback the student received	-	How does the student act upon received feedback?	Summative decision
21		Student's self-evaluation of performance or result Lecturer's evaluation of the student's performance or result	Discrepancy analysis of the student's self-evaluation and evaluation by the lecturer	How does the student self-regulate? What is the quality of the student's self-assessment?	Reflection assignment for the student based on the analysis results
22	Online discussion of presented student work	Student conversations	Length of silences	To what level of profoundness do students jointly develop knowledge?	Support Automated question / trigger / hint to

	Online collaboration and conversation				stimulate further conversation
23	Collaboration during group work	-	-	-	Guidance Summative decision
24	Collaboration during group work	-	Analysis of role and contribution of individual student in group assignment	What personal knowledge does the student develop during group work?	Support Student-lecturer conversation
25	Spontaneous interacting with peers	-	Support sought by the student Support given by the student Who contacts who Hand movement synchronicity during interaction: who imitates who Relation of student's final assignment characteristics to received peer feedback and/or peer's assignments	How does the student cooperate with peers in function of his/her individual learning process? How does cooperation during the learning process reflect in the final assessment?	Support Guide cooperation Refer to good practices of peers Explain difference between students
26	Providing online feedback to peer's work	Individual student's contributions to the discussion forum Peers' rating of the student's feedback	Analysis of feedback given by the student Peers' judgement of the student's feedback as useful	What is the quality of the student's peer feedback?	Summative decision

Note. The cells that are marked light grey are the first element the lecturer addressed describing the opportunity.